Started on	Thursday, 13 January 2022, 3:12 PM
State	Finished
Completed on	Thursday, 13 January 2022, 3:31 PM
Time taken	19 mins 20 secs
Marks	12.00/15.00
Grade	**24.00** out of 30.00 (**80**%)

Question **1**

Correct

Mark 1.00 out of 1.00

In data preprocessing, which of the following **is not** an objective of the *aggregation* of attributes

Select one:

a. Obtain a less detailed scale

b. Reduce the number of attributes or objects

c. Reduce the variability of data

d. Obtain a more detailed description of data

Question **2**

Correct

Mark 1.00 out of 1.00

Given the two binary vectors below, which is their similarity according to the Simple Matching Coefficient?

a b c d e f g h i j

1 0 0 0 1 0 1 1 0 1

1 0 1 1 1 0 1 0 1 0

Select one:

a. 0.5

b. 0.1

c. 0.2

d. 0.3

Question **3**

Incorrect

Mark 0.00 out of 1.00

Given the two binary vectors below, which is their similarity according to the Jaccard Coefficient?

a b c d e f g h i j

1 0 0 0 1 0 1 1 0 1

1 0 1 1 1 0 1 0 1 0

Select one:

a. 0.375

b. 0.1

c. 0.5

d. 0.2

Question **4**

Correct

Mark 1.00 out of 1.00

Given the definitions below:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

which of the formulas below computes the accuracy of a binary classifier?

Select one:

a. TP / (TP + FP)

b. TN / (TN + FP)

c. TP / (TP + FN)

d. (TP + TN) / (TP + FP + TN + FN)

Question **5**

Correct

Mark 1.00 out of 1.00

When developing a classifier, which of the following is a symptom of overfitting?

Select one:

a. The precision is much greater than the recall

b. The error rate in the test set is much smaller than the error rate in the training set

c. The error rate in the test set is much greater than the error rate in the training set

d. The error rate in the test set is more than 30%

Question **6**

Correct

Mark 1.00 out of 1.00

In a decision tree, an attribute which is used only in nodes near the leaves...

Select one:

a. ...has a high correlation with respect to the target

b. ...is irrelevant with respect to the target

c. ...guarantees high increment of purity

d. ...gives little insight with respect to the target

Question **7**

Incorrect

Mark 0.00 out of 1.00

When training a neural network, what is the *learning rate*?

a. A multiplying factor of the correction to be applied to the connection weights

b. The slope of the activation function in a specific node

c. The speed of convergence to a stable solution during the learning process

d. The ratio between the size of the hidden layer and the input layer of the network

Question **8**

Correct

Mark 1.00 out of 1.00

Which of the following is a strength of the clustering algorithm DBSCAN?

Select one or more:

a. Ability to find cluster with concavities

b. Very fast computation

c. Requires to set the number of clusters as a parameter

d. Ability to separate outliers from regular data

Question **9**

Correct

Mark 1.00 out of 1.00

What does K-means try to minimise?

Select one:

a. The *separation*, that is the sum of the squared distances of each cluster centroid with respect tho the global centroid of the dataset

b. The *separation*, that is the sum of the squared distances of each point with respect to its centroid

c. The *distortion*, that is the sum of the squared distances of each point with respect to its centroid

d. The *distortion*, that is the sum of the squared distances of each point with respect to the points of the other clusters

Question **10**

Correct

Mark 1.00 out of 1.00

Which of the statements below is true? (One or more)

Select one or more:

a. DBSCAN always stops to a configuration which gives the optimal number of clusters

b. Increasing the radius of the neighbourhood can decrease the number of noise points

c. DBSCAN can give good performance when clusters have concavities

d. Sometimes DBSCAN stops to a configuration which does not include any cluster

Question **11**

Correct

Mark 1.00 out of 1.00

Which of the statements below best describes the strategy of Apriori in finding the frequent itemsets?

Select one:

a. Evaluation of the support of the itemsets in an order such that the interesting parts of the search space are pruned as soon as possible

b. Evaluation of the support of the itemsets in an order such that uninteresting parts of the search space are pruned as soon as possible

c. Evaluation of the support of the itemsets in an order such that uninteresting parts of the search space are considered only at the end of the execution

d. Evaluation of the confidence of the itemsets in an order such that uninteresting parts of the search space are pruned as soon as possible

Question **12**

Incorrect

Mark 0.00 out of 1.00

Consider the transactional dataset below

ID	Items
1	A,B,C
2	A,B,D
3	B,D,E
4	C,D
5	A,C,D,E

Which is the *confidence* of the rule A,C ⇒ B?

Select one:

- a. 100%
- b. 40%
- c. 50%
- d. 20%

Question **13**

Correct

Mark 1.00 out of 1.00

In a dataset with D attributes, how many subsets of attributes should be considered for feature selection according to an exhaustive search?

Select one:

a. O(D!)

b. $O(D^2)$

c. $O(2^D)$

d. O(D)

Question **14**

Correct

Mark 1.00 out of 1.00

When is polynomial regression appropriate?

a. When it is necessary to project the data into a higher dimensional space

b. When there is more than one predicting attribute

c. When the relationship between the predicting variable and the target cannot be approximated as linear

d. When the target values are not linearly separable

Question **15**

Correct

Mark 1.00 out of 1.00

Which of the following *is not* a strength point of *Dbscan* with respect to *K-means*

Select one:

a. The *robustness* with respect to the number of attributes

b. The *effectiveness* even if there are clusters with non-convex shape

c. The *effectiveness*, even in presence of *noise*

d. The efficiency even in large datasets

Jump to...

Machine Learning - Python Lab ►